IMPH-0002

17/0.00-101-A
07 730452

PATENT

"EXPRESS MAIL" Mailing Label No. RB567179173
Date of Deposit JULY 15, 1991
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D. C. 20231
JO ANN EITHIAN
(MAILER)
(SIGNATURE)

5/81

ALLERGENIC PROTEINS AND PEPTIDES FROM JAPANESE CEDAR POLLEN

Field of the Invention

The present invention relates to the field of compositions and treatments for allergies. More particularly, the present invention relates to the field of compositions and treatments for allergy to Japanese cedar pollen.

Reference to Related Applications

The present application is a continuation of application Serial no. filed July 12, 1991 entitled ALLERGENIC PROTEINS AND PEPTIDES FROM JAPANESE CEDAR POLLEN in the names of Irwin J. Griffith and Joanne Pollock, the disclosures of which are hereby incorporated by reference.

Background of the Invention

Genetically predisposed individuals, who make up about 10% of the population, become hypersensitized (allergic) to antigens from a variety of environmental sources to which they are exposed. Those antigens that can induce immediate and/or delayed types of hypersensitivity are known as allergens. King, T.P., Adv. Immunol. 23: 77-105, (1976). Anaphylaxis or atopy, which includes the symptoms of hay fever, asthma, and hives, is one form of immediate allergy. It can be caused by a variety of atopic allergens, such as products of grasses, trees, weeds, animal dander, insects, and food, drugs, and chemicals.

The antibodies involved in atopic allergy belong primarily to the IgE class of immunoglobulins. IgE binds to mast cells and basophils. Upon combination of a specific

allergen with IgE bound to mast cells, the IgE is cross-linked on the cell surface, resulting in the physiological effects of IgE-antigen interaction. Degranulation results in release of, among other substances, histamine, heparin, a chemotactic factor for eosinophilic leukocytes and the leukotrienes, C4, D4, and E4, which cause prolonged constriction of bronchial smooth muscle cells. Hood, L.E. et al. Immunology (2nd ed.), The Benjamin/Cumming Publishing Co., Inc. (1984). These released substances are the mediators which result in allergic symptoms caused by a combination of IgE with a specific allergen. Through them, the effects of an allergen are manifested. Such effects may be systemic or local in nature, depending on the route by which the antigen entered the body and the pattern of deposition of IgE and mast cells. Local manifestations generally occur on epithelial surfaces at the location at which the allergen entered the body. Systemic effects can include anaphylaxis (anaphylactic shock), which is the result of an IgE-basophil response to circulating (intravascular) antigen.

Japanese cedar (Sugi; *Cryptomeria japonica*) pollinosis is one of the most important allergic diseases in Japan. The number of patients suffering from this disease is on the increase and in some areas, more than 10% of the population are affected. Treatment of Japanese cedar pollinosis by administration of intact cedar pollen to effect hyposensitization to the allergen has been attempted. Hyposensitization using intact cedar pollen, however, has drawbacks in that it can elicit anaphylaxis from the cedar pollen if high doses are used, whereas when low doses are used to avoid anaphylaxis, treatment must be continued for several years to build up a tolerance for the pollen.

The major allergen from Japanese cedar pollen has been purified and designated as Sugi basic protein (SBP) or *Cry j* I. This protein is reported to be a basic protein with a molecular weight of 41-50 kDa and a pI of 8.8. There appear to be multiple isoforms of the allergen, apparently due in part to differential glycosylation. Yasueda et al. (1983)

J. Allergy Clin. Immunol. 71: 77-86; and Taniai et al. (1988) FEBS Letters 239: 329-332. The sequence of the first twenty amino acids at the N-terminal end of *Cry j* I and a sixteen amino acid internal sequence have been determined. Taniai *supra*.

A second allergen from Japanese cedar pollen having a molecular weight of about 37 kDa known as *Cry j* II has also been reported. Sakaguchi et al. (1990) Allergy 45: 309-312. This allergen was found to have no immunological cross-reactivity with *Cry j* I. Most patients with Japanese cedar pollinosis were found to have IgE antibodies to both *Cry j* I and *Cry j* II, however, sera from some patients reacted with only *Cry j* I or *Cry j* II.

In addition to hyposensitization of Japanese cedar pollinosis patients with low doses of intact pollen, U.S. patent 4,939,239 issued July 3, 1990 to Matsuhashi et al. discloses a hyposensitization agent comprising a saccharide covalently linked to a Japanese cedar pollen allergen for hyposensitization of persons sensitive to Japanese cedar pollen. This hyposensitization agent is reported to enhance the production of IgG and IgM antibodies, but reduce production of IgE antibodies which are specific to the allergen and responsible for anaphylaxis and allergy. The allergens used in the hyposensitization agent preferably have an NH_2-terminal amino acid sequence of Asp-Asn-Pro-Ile-Asp-Ser-X-Trp-Arg-Gly-Asp-Ser-Asn-Trp-Ala-Gln-Asn-Arg-Met-Lys-, wherein X is Ser, Cys, Thr, or His (SEQ ID NO: 18). Additionally, Usui et al. (1990) Int. Arch. Allergy Appl. Immunol. 91: 74-79 reported that the ability of a Sugi basic protein (i.e., *Cry j* I)-pullulan conjugate to elicit the Arthus reaction was markedly reduced, about 1,000 times lower than that of native Sugi basic protein and suggested that the Sugi basic protein-pullulan conjugate would be a good candidate for desensitization therapy against cedar pollinosis.

The *Cry j* I allergen found in *C. japonica* has also been found to be cross-reactive with allergens in the pollen

from other species of trees, including *Cupressus sempervirens*. Panzani et al. (Annals of Allergy 57: 26-30 (1986) reported that cross reactivity was detected between allergens in the pollens of *Cupressus sempervirens* and *Cryptomeria japonica* in skin testing, RAST and RAST inhibition. A 50 kDa allergen isolated from Mountain Cedar (*Juniperus sabinoides*) which has the NH_2-terminal sequence AspAsnProIleAsp (SEQ ID NO: 25) (Gross et al, (1978) Scand. J. Immunol. 8: 437-441) which is the same sequence as the first five amino acids of the NH_{-2} terminal end of the *Cry j* I allergen.

Despite the attention Japanese cedar pollinosis allergens have received, definition or characterization of the allergens responsible for its adverse effects on people is far from complete. Current desensitization therapy involves treatment with intact pollen, with its attendant risks of anaphylaxis if high doses of pollen are administered, or long desensitization times when low doses of pollen are administered.

Summary of the Invention

The present invention provides nucleic acid sequences coding for the *Cryptomeria japonica* major pollen allergen *Cry j* I and fragments thereof. The present invention also provides purified *Cry j* I and at least one fragment thereof produced in a host cell transformed with a nucleic acid sequence coding for *Cry j* I or at least one fragment thereof and fragments of *Cry j* prepared synthetically. *Cry j* I and fragments thereof are useful for diagnosing, treating, and preventing Japanese cedar pollinosis. This invention is more particularly described in the appended claims and is described in its preferred embodiments in the following description.

Brief Description of the Drawings

Figures 1a and 1b show the composite nucleic acid sequence from the two overlapping clones JC 71.6 and pUC19J91A coding for *Cry j* I. The complete cDNA sequence for *Cry j* I is composed of 1312 nucleotides, including 66 nucleotides of

5' untranslated sequence, an open reading frame starting with the codon for an initiating methionine, of 1122 nucleotides, and a 3' untranslated region. Figures 1a and 1b also show the deduced amino acid sequence of *Cry j* I.

Detailed Description of the Invention

The present invention provides nucleic acid sequences coding for *Cry j* I, the major allergen found in Japanese cedar pollen. The nucleic acid sequence coding for *Cry j* I preferably has the sequence shown in Figures 1a and 1b (SEQ ID NO: 1). The nucleic acid sequence coding for *Cry j* I shown in Figures 1a and 1b (SEQ ID NO: 1) contains a 21 amino acid leader sequence from base 66 through base 128. This leader sequence is cleaved from the mature protein which is encoded by bases 129 through 1187. The deduced amino acid sequence of *Cry j* I is also shown in Figures 1a and 1b (SEQ ID NO: 2). The nucleic acid sequence of the invention codes for a protein having a predicted molecular weight of 38.5 kDa, with a pI of 7.8, and five potential N-linked glycosylation sites. Utilization of these glycosylation sites will increase the molecular weight and affect the pI of the mature protein. The deduced amino acid sequence for the mature protein encoded by the nucleic acid sequence of the invention is identical with the known NH_2-terminal and internal amino acid sequences reported by Taniai et al., *supra*. The NH_2-terminal end of *Cry j* I reported by Tanai et al., *supra* has the sequence shown in SEQ ID NO: 18. The internal sequence reported by Taniai et al., *supra* has the sequence GluAlaPheAsnValGluAsnGlyAsnAlaThrProGlnLeuThrLys (SEQ ID NO: 19). There is a possible minor sequence polymorphism observed in the nucleic acid sequence of the invention. For example, a single nucleotide substitution in the codon for amino acid 74 (GGG ↔ GAG) has been detected in different clones that may result in amino acid polymorphism (G vs. E) at this site. It is expected that there are additional sequence polymorphisms, and it will be appreciated by one skilled in the art that one or more nucleotides (up to about 1% of the nucleotides) in the

nucleic acid sequence coding for *Cry j* I may vary among individual *C. japonica* plants due to natural allelic variation. Any and all such nucleotide variations and resulting amino acid polymorphisms are within the scope of the invention.

Fragments of the nucleic acid sequence coding for *Cry j* I are also within the scope of the invention. Fragments within the scope of the invention include those coding for parts of *Cry j* I that elicit an antigenic response in mammals, preferably humans, such as the stimulation of minimal amounts of IgE; the eliciting of IgG and IgM antibodies; or the eliciting of a T cell response such as proliferation and/or lymphokine secretion and/or the induction of T cell anergy. The foregoing fragments of *Cry j* I are referred to herein as antigenic fragments. Fragments within the scope of the invention also include those capable of hybridizing with nucleic acid from other plant species for use in screening protocols to detect allergens that are cross-reactive with *Cry j* I. As used herein, a fragment of the nucleic acid sequence coding for *Cry j* I refers to a nucleotide sequence having fewer bases than the nucleotide sequence coding for the entire amino acid sequence of *Cry j* I and/or mature *Cry j* I. Generally, the nucleic acid sequence coding for the fragment or fragments of *Cry j* I will be selected from the bases coding for the mature protein, however, in some instances it may be desirable to select all or a part of a fragment or fragments from the leader sequence portion of the nucleic acid sequence of the invention. The nucleic acid sequence of the invention may also contain linker sequences, restriction endonuclease sites and other sequences useful for cloning, expression or purification of *Cry j* I or fragments thereof.

A nucleic acid sequence coding for *Cry j* I may be obtained from any part of *C. japonica* plants. However, Applicants have found that mRNA coding for *Cry j* I could not be obtained from commercially available *C. japonica* pollen. This inability to obtain mRNA from the pollen may be due to problems with storage or transportation of commercially

available pollen. Applicants have found that fresh pollen and staminate cones are a good source of *Cry j* I mRNA. It may also be possible to obtain the nucleic acid sequence coding for *Cry j* I from genomic DNA. *C. japonica* is a well-known species of cedar, and plant material may be obtained from wild, cultivated, or ornamental plants. The nucleic acid sequence coding for *Cry j* I may be obtained using the method disclosed herein or any other suitable techniques for isolation and cloning of genes. The nucleic acid sequence of the invention may be DNA or RNA.

The present invention provides expression vectors and host cells transformed to express the nucleic acid sequences of the invention. Nucleic acid coding for *Cry j* I, or at least one fragment thereof may be expressed in bacterial cells such as *E. coli*, insect cells, yeast, or mammalian cells such as Chinese hamster ovary cells (CHO). Suitable expression vectors, promoters, enhancers, and other expression control elements may be found in Sambrook et al. *Molecular Cloning: A Laboratory Manual*, second edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989. Expression in yeast, insect or mammalian cells would lead to partial or complete glycosylation of the recombinant material and formation of any inter- or intra-chain disulfide bonds, if such exist. Suitable vectors for expression in yeast include YepSec1 (Baldari et al. (1987) Embo J. <u>6</u>: 229-234); pMFα (Kurjan and Herskowitz (1982) Cell <u>30</u>: 933-943); and JRY88 (Schultz et al. (1987) Gene <u>54</u>: 113-123).

For expression in *E. coli*, suitable expression vectors include pTRC (Amann et al. (1988) Gene <u>69</u>: 301-315); pGEX (Amrad Corp., Melbourne, Australia); pMAL (N.E. Biolabs, Beverly, MA); pRIT5 (Pharmacia, Piscataway, NJ); and pSEM (Knapp et al. (1990) BioTechniques <u>8</u>: 280-281). The use of pTRC and pGEX will lead to the expression of unfused protein. The use of pMAL, pRIT5 and pSEM will lead to the expression of allergen fused to maltose E binding protein (pMAL), protein A (pRIT5), or truncated β-galactosidase (PSEM). When *Cry j* I, fragment, or fragments thereof is expressed as a fusion

protein, it is particularly advantageous to introduce an enzymatic cleavage site at the fusion junction between the carrier protein and *Cry j* I or fragment thereof. *Cry j* I or fragment thereof may then be recovered from the fusion protein through enzymatic cleavage at the enzymatic site and biochemical purification using conventional techniques for purification of proteins and peptides. Suitable enzymatic cleavage sites include those for blood clotting Factor X or thrombin for which the appropriate enzymes and protocols for cleavage are commercially available from for example Sigma Chemical Company, St. Louis, MO and N.E. Biolabs, Beverly, MA.

Host cells can be transformed to express the nucleic acid sequences of the invention using conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, or electroporation. Suitable methods for transforming the host cells may be found in Sambrook et al. *supra*, and other laboratory textbooks.

The nucleic acid sequences of the invention may also be synthesized using standard techniques.

The present invention also provides a method producing Japanese cedar pollen allergen *Cry j* I or at least one fragment thereof comprising the steps of culturing a host cell transformed with a DNA sequence encoding Japanese cedar pollen allergen *Cry j* I or at least one fragment thereof in an appropriate medium to produce a mixture of cells and medium containing said Japanese cedar pollen allergen *Cry j* I or at least one fragment thereof; and purifying the mixture to produce substantially pure Japanese cedar pollen allergen *Cry j* I or at least one fragment thereof. Host cells transformed with an expression vector containing DNA coding for *Cry j* I or at least one fragment thereof are cultured in a suitable medium for the host cell. *Cry j* I protein and peptides can be purified from cell culture medium, host cells, or both using techniques known in the art for purifying peptides and proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis and

immunopurification with antibodies specific for *Cry j* I or fragments thereof. *Cry j* I protein and fragments thereof can also be purified using the procedures in Yasueda et al., *supra*, or U.S. patent 4,939,239. The terms isolated and purified are used interchangeably herein and refer to peptides, protein, protein fragments, and nucleic acid sequences substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when synthesized chemically.

Another aspect of the invention provides preparations comprising Japanese cedar pollen allergen *Cry j* I or at least one fragment thereof synthesized in a host cell transformed with a DNA sequence encoding all or a portion of Japanese cedar pollen allergen *Cry j* I, or chemically synthesized, and purified Japanese cedar pollen allergen *Cry j* I protein, or at least one antigenic fragment thereof produced in a host cell transformed with a nucleic acid sequence of the invention, or chemically synthesized. In preferred embodiments of the invention the *Cry j* I protein is produced in a host cell transformed with the nucleic acid sequence coding for at least the mature *Cry j* I protein.

Fragments of an allergen from Japanese cedar pollen, preferably *Cry j* I, eliciting a desired antigenic response (referred to herein as antigenic fragments) may be obtained, for example, by screening peptides synthesized from the corresponding fragment of the nucleic acid sequence of the invention coding for such peptides or synthesized chemically using techniques known in the art. The peptide fragments of the allergen may be obtained by any method known in the art such as chemical cleavage of the allergen, arbitrary division of the allergen into fragments of a desired length with no overlap of the peptides, or preferably division of the allergen into overlapping fragments of a desired length. The fragments are tested to determine their antigenicity and allergenicity. Fragments of *Cry j* I which are capable of eliciting a T cell response such as stimulation (i.e.,

proliferation or lymphokine secretion) and/or are capable of inducing T cell anergy are particularly desirable. Fragments of *Cry j* I which do not bind immunoglobulin E (IgE) and/or which have minimal IgE stimulating activity are also desirable. If the fragment or fragments of *Cry j* I bind IgE, it is preferable that such binding does not lead to histamine release, e.g., such binding does not cause cross-linking of IgE on mast cells. Minimal IgE stimulating activity refers to IgE stimulating activity that is less than the amount of IgE production stimulated by the whole *Cry j* I protein. Preferred fragments also include antigenic fragments which, when administered to a Japanese cedar pollen-sensitive individual, are capable of modifying the allergic response to Japanese cedar pollen of the individual, and antigenic fragments which, when administered to a Japanese cedar pollen-sensitive individual, are capable of modifying B-cell response, T-cell response or both B-cell and T-cell response of the individual to a Japanese cedar pollen antigen.

The *Cry j* protein or fragments thereof are preferably tested in mammalian models of Japanese cedar pollinosis such as the mouse model disclosed in Tamura et al. (1986) Microbiol. Immunol. 30: 883-896, or U.S. patent 4,939,239; or the primate model disclosed in Chiba et al. (1990) Int. Arch. Allergy Immunol. 93: 83-88. Initial screening for IgE binding to the protein or fragments thereof may be performed by scratch tests or intradermal skin tests on laboratory animals or human volunteers, or in *in vitro* systems such as RAST (radioallergosorbent test), RAST inhibition, ELISA assay or radioimmunoassay (RIA).

The *Cry j* I protein, and fragments or portions derived therefrom (peptides) can be used in methods of diagnosing, treating and preventing allergic reactions to Japanese cedar pollen. Thus the present invention provides therapeutic compositions comprising purified Japanese cedar pollen allergen *Cry j* I or at least one fragment thereof produced in a host cell transformed to express *Cry j* I or at least one fragment thereof, and a pharmaceutically acceptable

carrier or diluent. The therapeutic compositions of the invention may also comprise synthetically prepared *Cry j* I or at least one fragment thereof and a pharmaceutically acceptable carrier or diluent. Administration of the therapeutic compositions of the present invention to an individual to be desensitized can be carried out using known techniques. *Cry j* I protein or at least one fragment thereof may be administered to an individual in combination with, for example, an appropriate diluent, a carrier and/or an adjuvant. Pharmaceutically acceptable diluents include saline and aqueous buffer solutions. Pharmaceutically acceptable carriers include polyethylene glycol (Wie et al. (1981) Int. Arch. Allergy Appl. Immunol. 64:84-99) and liposomes (Strejan et al. (1984) J Neuroimmunol 7: 27). Such compositions will generally be administered by injection, oral administration, inhalation, transdermal application or rectal administration. The therapeutic compositions of the invention are administered to Japanese cedar pollen-sensitive individuals at dosages and for lengths of time effective to reduce sensitivity (i.e, reduce the allergic response) of the individual to Japanese cedar pollen. Effective amounts of the therapeutic compositions will vary according to factors such as the degree of sensitivity of the individual to Japanese cedar pollen, the age, sex, and weight of the individual, and the ability of the *Cry j* I protein or fragment thereof to elicit an antigenic response in the individual.

The *Cry j* I cDNA (or the mRNA from which it was transcribed) or a portion thereof can be used to identify similar sequences in any variety or type of plant and thus, to identify or "pull out" sequences which have sufficient homology to hybridize to the *Cry j* I cDNA or mRNA or portion thereof, for example, DNA from *Cupressus sempervirens* or *Juniperus sabinoides*, under conditions of low stringency. Those sequences which have sufficient homology (generally greater than 40%) can be selected for further assessment using the method described herein. Alternatively, high stringency conditions can be used. In this manner, DNA of the present

invention can be used to identify, in other types of plants, preferably related families, genera, or species such as *Juniperus*, or *Cupressus*, sequences encoding polypeptides having amino acid sequences similar to that of Japanese cedar pollen allergen *Cry j I*, and thus to identify allergens in other species. Thus, the present invention includes not only *Cry j I*, but also other allergens encoded by DNA which hybridizes to DNA of the present invention. The invention further includes isolated allergenic proteins or fragments thereof that are immunologically related to *Cry j* I or fragments thereof, such as by antibody cross-reactivity wherein the isolated allergenic proteins or fragments thereof are capable of binding to antibodies specific for the protein and peptides of the invention, etc.

Proteins or peptides encoded by the cDNA of the present invention can be used, for example as "purified" allergens. Such purified allergens are useful in the standardization of allergen extracts which are key reagents for the diagnosis and treatment of Japanese cedar pollinosis. Furthermore, by using peptides based on the nucleic acid sequences of *Cry j* I, anti-peptide antisera or monoclonal antibodies can be made using standard methods. These sera or monoclonal antibodies can be used to standardize allergen extracts.

Through use of the peptides and protein of the present invention, preparations of consistent, well-defined composition and biological activity can be made and administered for therapeutic purposes (e.g. to modify the allergic response of a Japanese cedar sensitive individual to pollen of such trees). Administration of such peptides or protein may, for example, modify B-cell response to *Cry j* I allergen, T-cell response to *Cry j* I allergen or both responses. Purified peptides can also be used to study the mechanism of immunotherapy of *C. japonica* allergy and to design modified derivatives or analogues useful in immunotherapy.

Work by others has shown that high doses of

allergens generally produce the best results (i.e., best symptom relief). However, many people are unable to tolerate large doses of allergens because of allergic reactions to the allergens. Modification of naturally-occurring allergens can be designed in such a manner that modified peptides or modified allergens which have the same or enhanced therapeutic properties as the corresponding naturally-occurring allergen but have reduced side effects (especially anaphylactic reactions) can be produced. These can be, for example, a protein or peptide of the present invention (e.g., one having all or a portion of the amino acid sequence of *Cry j* I), or a modified protein or peptide, or protein or peptide analogue (e.g., a protein or peptide in which the amino acid sequence has been altered, such as by amino acid substitution, deletion, or addition, to modify immunogenicity and/or reduce allergenicity or to which a component has been added for the same purpose). For example, *Cry j* I protein or peptides can be modified using the polyethylene glycol method of A. Sehon and co-workers. Wie et al., *supra*.

Modification of *Cry j* I protein or peptides can also include reduction/alkylation (Tarr [1986] in: Methods of Protein Microcharacterization, J.E. Silver, ed. Humana Press, Clifton, NJ, pp 155-194); acylation (Tarr, *supra*); esterification (Tarr, *supra*); chemical coupling to an appropriate carrier (Mishell and Shiigi, eds, [1980] Selected Methods in Cellular Immunology, WH Freeman, San Francisco, CA; U.S. patent 4,939,239); or mild formalin treatment (Marsh [1971] Int. Arch. Allergy Appl. Immunol. 41: 199-215).

Using the structural information now available, it is possible to design *Cry j* I peptides which, when administered to a Japanese cedar pollen sensitive individual in sufficient quantities, will modify the individual's allergic response to Japanese cedar pollen. This can be done, for example, by examining the structure of *Cry j* I, producing peptides (via an expression system, synthetically or otherwise) to be examined for their ability to influence B-cell and/or T-cell responses in Japanese cedar pollen

sensitive individuals and selecting appropriate epitopes recognized by the cells. In referring to an epitope, the epitope will be the basic element or smallest unit of recognition by a receptor, particularly immunoglobulins, histocompatibility antigens and T cell receptors where the amino acids essential to the receptor recognition may be contiguous and/or non-contiguous in the amino acid sequence. Amino acid sequences which mimic those of the epitopes and which are capable of down regulating allergic response to *Cry j* I can also be used.

It is now also possible to design an agent or a drug capable of blocking or inhibiting the ability of Japanese cedar pollen allergen to induce an allergic reaction in Japanese cedar pollen sensitive individuals. Such agents could be designed, for example, in such a manner that they would bind to relevant anti-*Cry j* I-IgE's, thus preventing IgE-allergen binding and subsequent mast cell degranulation. Alternatively, such agents could bind to cellular components of the immune system, resulting in suppression or desensitization of the allergic response to *C. japonica* pollen allergens. A non-restrictive example of this is the use of appropriate B- and T-cell epitope peptides, or modifications thereof, based on the cDNA/protein structures of the present invention to suppress the allergic response to Japanese cedar pollen. This can be carried out by defining the structures of B- and T-cell epitope peptides which affect B- and T-cell function in *in vitro* studies with blood components from Japanese cedar pollen sensitive individuals.

Protein, peptides or antibodies of the present invention can also be used for detecting and diagnosing Japanese cedar pollinosis. For example, this could be done by combining blood or blood products obtained from an individual to be assessed for sensitivity to Japanese cedar pollen with an isolated antigenic peptide or peptides of *Cry j* I, or isolated *Cry j* I protein, under conditions appropriate for binding of components (e.g., antibodies, T-cells, B-cells) in the blood with the peptide(s) or protein and

determining the extent to which such binding occurs.

The present invention also provides a method of producing *Cry j* I or fragment thereof comprising culturing a host cell containing an expression vector which contains DNA encoding all or a fragment of *Cry j* I under conditions appropriate for expression of *Cry j* I or fragment. The expressed product is then recovered, using known techniques. Alternatively, *Cry j* I or fragment thereof can be synthesized using known mechanical or chemical techniques.

The DNA used in any embodiment of this invention can be cDNA obtained as described herein, or alternatively, can be any oligodeoxynucleotide sequence having all or a portion of a sequence represented herein, or their functional equivalents. Such oligodeoxynucleotide sequences can be produced chemically or mechanically, using known techniques. A functional equivalent of an oligonucleotide sequence is one which is 1) a sequence capable of hybridizing to a complementary oligonucleotide to which the sequence (or corresponding sequence portions) of SEQ ID NO: 1 or fragments thereof hybridizes, or 2) the sequence (or corresponding sequence portion) complementary to SEQ ID NO: 1, and/or 3) a sequence which encodes a product (e.g., a polypeptide or peptide) having the same functional characteristics of the product encoded by the sequence (or corresponding sequence portion) of SEQ ID NO: 1. Whether a functional equivalent must meet one or both criteria will depend on its use (e.g., if it is to be used only as an oligoprobe, it need meet only the first or second criteria and if it is to be used to produce a *Cry j* I allergen, it need only meet the third criterion).

Experimental

Example 1

Attempted Extraction of RNA From Japanese Cedar Pollen

Multiple attempts were made to obtain RNA from commercially-available, non-defatted, *Cryptomeria japonica* (Japanese cedar) pollen (Hollister Stier, Seattle, WA). Initially, the method of Sambrook et al., *Molecular Cloning.*

A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989 was used in which the sample was suspended and lysed in 4 M guanidine buffer, ground under liquid nitrogen, and pelleted through 5.7 M cesium chloride by ultracentrifugation. Various amounts (3, 5 and 10 g) of pollen in varying amounts of guanidine lysis buffer (10 and 25 ml) were tried. Centrifugation through cesium resulted in viscous material in the bottom of the tube, from which it was not possible to recover an RNA pellet. Although it was possible to obtain RNA from defatted *Ambrosia artemisiifolia* (ragweed) pollen (Greer Laboratories, Lenior, NC) using this protocol, defatting the *Cryptomeria japonica* pollen with acetone before guanidine extraction also did not yield any RNA, as determined by absorbance at A_{260}.

An acid phenol extraction of RNA according to the method in *Sambrook* et al., *supra* was attempted from *Cryptomeria japonica* pollen. The pollen was ground and sheared in 4.5 M guanidine solution, acidified by addition of 2 M sodium acetate, and extracted with water-saturated phenol plus chloroform. After precipitation, the pellet was washed with 4 M lithium chloride, redissolved in 10 mM Tris/5 mM EDTA/1% SDS, chloroform extracted, and re-precipitated with NaCl and absolute ethanol. It was possible to extract *Ambrosia artemisiifolia* but not *Cryptomeria japonica* RNA with this procedure.

Next, 4 g of *Cryptomeria japonica* pollen was suspended in 10 ml extraction buffer (50 mM Tris, pH 9.0, 0.2 M NaCl, 10 mM Mg acetate and diethylpyrocarbonate (DEPC) to 0.1%), ground in a mortar and pestle on dry ice, transferred to a centrifuge tube with 1% SDS, 10 mM EDTA and 0.5% N-lauroyl sarcosine, and the mixture was extracted five times with warm phenol. The aqueous phase was recovered after the final centrifugation, 2.5 vol. absolute ethanol was added, and the mixture was incubated overnight at 4°C. The pellet was recovered by centrifugation, resuspended in 1 ml dH20 by heating to 65°C, and reprecipitated by the addition of 0.1 vol. 3 M Na acetate and 2.0 vol. of ethanol. No detectable

RNA was recovered in the pellet as judged by absorbance at A_{260} and gel electrophoresis.

Finally, 500 mg of *Cryptomeria japonica* pollen was ground by mortar and pestle on dry ice and suspended in 5 ml of 50 mM Tris pH 9.0 with 0.2 M NaCl, 1 mM EDTA, 1% SDS that had been treated overnight with 0.1% DEPC, as previously described in Frankis and Mascarhenas (1980) Ann. Bot. 45: 595-599. After five extractions with phenol/chloroform/isoamyl alcohol (mixed at 25:24:1), material was precipitated from the aqueous phase with 0.1 volume 3 M sodium acetate and 2 volumes ethanol. The pellet was recovered by centrifugation, resuspended in dH_2O and heated to 65°C to solubilize the precipitated material. Further precipitations with lithium chloride were not done. There was no detectable RNA recovered, as determined by absorbance at A_{260} and gel electrophoresis.

In summary, it has not been possible to recover RNA from the commercial pollen. It is not known whether the RNA has been degraded during storage or shipment, or whether the protocols used in this example did not allow recovery of extant RNA. However, RNA was recovered from fresh *Cryptomeria japonica* pollen and staminate cone samples. (See Example 2.)

Example 2

Extraction of RNA From Japanese Cedar Pollen and Staminate Cones and Cloning of *Cry j I*

Fresh pollen and staminate cone samples, collected from a single *Cryptomeria japonica* (Japanese cedar) tree at the Arnold Arboretum (Boston, MA), were frozen immediately on dry ice. RNA was prepared from 500 mg of each sample, essentially as described by Frankis and Mascarenhas, *supra*. The samples were ground by mortar and pestle on dry ice and suspended in 5 ml of 50 mM Tris pH 9.0 with 0.2 M NaCl, 1 mM EDTA, 1% SDS that had been treated overnight with 0.1% DEPC. After five extractions with phenol/chloroform/isoamyl alcohol (mixed at 25:24:1), the RNA was precipitated from the aqueous phase with 0.1 volume 2 M sodium acetate and 2 volumes ethanol. The pellets were recovered by centrifugation,

resuspended in dH_2O and heated to 65°C for 5 min. Two ml of 4 M lithium chloride were added to the RNA preparations and they were incubated overnight at 0°C. The RNA pellets were recovered by centrifugation, resuspended in 1 ml dH_2O, and again precipitated with 3 M sodium acetate and ethanol overnight. The final pellets were resuspended in 100 μl dH_2O and stored at -80°C.

First strand cDNA was synthesized from 8 μg flowerhead and 4 μg pollen RNA using a commercially available kit (cDNA synthesis systems kit, BRL, Gaithersburg, MD) with oligo dT priming according to the method of Gubler and Hoffman (1983) Gene 25: 263-269. An attempt was made to amplify cDNA encoding *Cry j* I using the degenerate oligonucleotide CP-1 (which has the sequence 5'-GATAATCCGATAGATAG-3', wherein T at position 3 can also be C; T at position 6 can also be C; G at position 9 can also be A,T, or C; A at position 12 can also be T, or C; T at position 15 can also be C; A at position 16 can also be T; and G at position 17 can also be C; SEQ ID NO: 3) and primers EDT and ED. Primer EDT has the sequence 5'GGAATTCTCTAGACTGCAGGTTTTTTTTTTTTTTT-3'(SEQ ID NO: 24). Primer ED has the sequence 5'-GGAATTCTCTAGACTGCAGGT-3' (SEQ ID NO: 23). CP-1 is the degenerate oligonucleotide sequence encoding the first six amino acids of the amino terminus (AspAsnProIleAspSer, amino acids 1-6 of SEQ ID NO: 1) of *Cry j* I. EDT will hybridize with the poly A tail of the gene. All oligonucleotides were synthesized by Research Genetics, Inc. Huntsville, AL. Polymerase chain reactions (PCR) were carried out using a commercially available kit(GeneAmp DNA Amplification kit, Perkin Elmer Cetus, Norwalk, CT) whereby 10 μl 10x buffer containing dNTPs was mixed with 1μg of CP-1 and 1 μg of ED/EDT primers (ED:EDT in a 3:1 M ratio), cDNA (3-5 μl of a 20 μl first strand cDNA reaction mix), 0.5 μl Amplitaq DNA polymerase, and distilled water to 100 μl.

The samples were amplified with a programmable thermal controller (MJ Research, Inc., Cambridge, MA). The first 5 rounds of amplification consisted of denaturation at 94°C for 1 minute, annealing of primers to the template at

45°C for 1.5 minutes, and chain elongation at 70°C for 2 minutes. The final 20 rounds of amplification consisted of denaturation as above, annealing at 55°C for 1.5 minutes, and elongation as above. Five percent (5 μl) of this initial amplification was then used in a secondary amplification with 1 μg each of CP-2 (which has the sequence 5'-GGGAATTCAATTGGGCGCAGAATGG-3' wherein T at position 11 can also be C; G at position 17 can also be A, T, or C; G at position 20 can also be A; T at position 23 can also be C; and G at position 24 can also be C) (SEQ ID NO: 4), a nested primer, and ED, as above. The sequence 5'-GGGAATTC-3' (bases 1 through 8 of SEQ ID NO: 4) in primer CP-2 represents an *Eco* R1 site added for cloning purposes; the remaining degenerate oligonucleotide sequence encodes amino acids 13-18 of *Cry j* I (AsnTrpAlaGlnAsnArg, amino acids 13 through 18 of SEQ ID NO: 1). Multiple DNA bands were resolved on a 1% GTG agarose gel (FMC, Rockport, ME), none of which hybridized with ^{32}P end-labeled probe CP-3 (SEQ ID NO: 5) in a Southern blot performed according to the method in Sambrook et al. *supra*. Therefore, it was not possible to select a specific *Cry j* I DNA band and this approach was not pursued. CP-3 has the sequence 5'-CTGCAGCCATTTTCIACATTAAA-3' wherein A at position 9 can also be G; T at position 12 can also be C; A at position 18 can also be G; and A at position 21 can also be G) (SEQ ID NO: 5). Inosine (I) is used at position 15 in place of G or A or T or C to reduce degeneracy (Knoth et al. (1988) Nucleic Acids Res. 16: 10932). The sequence 5'-CTGCAG-3' (bases 1 through 6 of SEQ ID NO: 5) in primer CP-3 represent a Pst I site added for cloning purposes; the remaining degenerate oligonucleotide sequence is the non-coding strand sequence corresponding to coding strand sequence encoding amino acids PheAsnValGluAsnGly (amino acids 327 through 332 of SEQ ID NO: 1) from the internal sequence of *Cry j* I.

A primary PCR was also performed on first-strand cDNA using CP-1 (SEQ ID NO: 3) and CP-3 (SEQ ID NO: 5), as above. A secondary PCR was performed using 5% of the primary reaction using CP-2 (SEQ ID NO: 4) and CP-3 (SEQ ID NO: 5).

Again, multiple bands were observed, none of which could be specifically identified in a Southern blot as *Cry j* I, and this approach was also not pursued.

Double-stranded cDNA was then synthesized from approximately 4 μg (pollen) or 8 μg (flowerhead) RNA using a commercially available kit (cDNA synthesis system plus kit, BRL, Gaithersburg, MD). After a phenol extraction and ethanol precipitation, the cDNA was blunted with T4 DNA polymerase (Promega, Madison, WI), and ligated to ethanol precipitated, self-annealed, AT (SEQ ID NO: 20) and AL (SEQ ID NO: 22) oligonucleotides for use in a modified Anchored PCR reaction, according to the method in Rafnar et al. (1991) J. Biol. Chem. 266: 1229-1236; Frohman et al. (1990) Proc. Natl. Acad. Sci. USA 85: 8998-9002; and Roux et al. (1990) BioTech. 8: 48-57. Oligonucleotide AT has the sequence 5'-GGGTCTAGAGGTACCGTCCGATCGATCATT-3'(SEQ ID NO: 20) (Rafnar et al. *supra*). Oligonucleotide AL has the sequence AATGATCGATGCT (SEQ ID NO: 22) (Rafnar et al. *Supra*. The amino terminus of *Cry j* I was amplified from the linkered cDNA (3 μl from a 20 μl reaction) with 1 μg each of oligonucleotides AP (SEQ ID NO: 21) and degenerate *Cry j* I primer CP-7 (which has the sequence 5'-TTCATICGATTCTGGGCCCA-3' wherein G at position 8 can also be T; A at position 9 can also be G; C at position 12 can also be T; and G at position 15 can also be A, T, or C)(SEQ ID NO: 6). Inosine (I) is used at position 6 in place of G or A or T or C to reduce degeneracy (Knoth et al. *supra*). The degenerate oligonucleotide CP-7 (SEQ ID NO: 6) is the non-coding strand sequence corresponding to coding strand sequence encoding amino acids 14-20 (TrpAlaGlnAsnArgMetLys) from the amino terminus of *Cry j* I (amino acids 14-20 of SEQ ID NO: 1). Oligonucleotide AP has the sequence 5'-GGGTCTAGAGGTACCGTCCG-3' (SEQ ID NO: 21).

The primary PCR reaction was carried out as described herein. Five percent (5 μl) of this initial amplification was then used in a secondary amplification with 1 μg each of AP (SEQ ID NO: 21) and degenerate *Cry j* I primer CP-8 (SEQ ID NO: 7) an internally nested *Cry j* I

oligonucleotide primer, as described herein. Primer CP-8 has the sequence 5'-CCTGCAGCGATTCTGGGCCCAAATT-3' wherein G at position 9 can also be T; A at position 10 can also be G; C at position 13 can also be T; G at position 16 can also be A, T, or C; and A at position 23 can also be G)(SEQ ID NO: 7). The nucleotides 5'-CCTGCAG-3' (bases 1 through 7 of SEQ ID NO: 7) represent a *Pst* I restriction site added for cloning purposes. The remaining degenerate oligonucleotide sequence is the non-coding strand sequence corresponding to coding strand sequence encoding amino acids 13-18 of *Cry j* I (AsnTrpAlaGlnAsnArg, amino acids 13-18 of SEQ ID NO: 1) from the amino terminus of *Cry j* I. The dominant amplified product was a DNA band of approximately 193 base pairs, as visualized on an ethidium bromide (EtBr)-stained 3% GTG agarose gel.

Amplified DNA was recovered by sequential chloroform, phenol, and chloroform extractions, followed by precipitation at -20°C with 0.5 volumes of 7.5 ammonium acetate and 1.5 volumes of isopropanol. After precipitation and washing with 70% ethanol, the DNA was simultaneously digested with *Xba* I and *Pst* I in a 15 μl reaction and electrophoresed through a preparative 3% GTG NuSieve low melt gel (FMC, Rockport, ME). The appropriate sized DNA band was visualized by EtBr staining, excised, and ligated into appropriately digested M13mp18 for sequencing by the dideoxy chain termination method (Sanger et al. (1977) Proc. Natl Acad Sci USA 74: 5463-5476) using a commercially available sequencing kit (sequenase kit, U.S. Biochemicals, Cleveland, OH). It was initially thought that ligatable material could only be derived from staminate cone-derived RNA. However, upon subsequent examination, it was shown that ligatable material could be recovered from PCR product generated from pollen-derived RNA, and from staminate cone-derived RNA.

The clone designated JC71.6 was found to contain a partial sequence of *Cry j* I. This was confirmed as an authentic clone of *Cry j* I by having complete identity to the disclosed NH_2-terminal sequence of *Cry j* I (Taniai et al.

supra). The amino acid at position 7 was determined to be cysteine (Cys) in agreement with the sequence disclosed in U.S. patent 4, 939,239. Amino acid numbering is based on the sequence of the mature protein; amino acid 1 corresponds to the aspartic acid (Asp) disclosed as the NH_2-terminus of *Cry j* I (Taniai et al. *supra*) The initiating methionine was found to be amino acid -21 relative to the first amino acid of the mature protein. The position of the initiating methionine was supported by the presence of upstream in-frame- stop codons and by 78 % homology of the surrounding nucleotide sequence with the plant consensus sequence that encompasses the initiating methionine, as reported by Lutcke et al. (1987) EMBO J. 6: 43-48.

The cDNA encoding the remainder of *Cry j* I gene was cloned from the linkered cDNA by using oligonucleotides CP-9 (which has the sequence 5'-ATGGATTCCCCTTGCTTA-3')(SEQ ID NO: 8) and AP (SEQ ID NO: 21) in the primary PCR reaction. Oligonucleotide CP-9 (SEQ ID NO: 8) encodes amino acids MetAspSerProCysLeu of *Cry j* I (amino acids -21 through -16 of SEQ ID NO: 1) from the leader sequence of *Cry j* I, and is based on the nucleotide sequence determined for the partial *Cry j* I clone JC76.1.

A secondary PCR reaction was performed on 5% of the initial amplification mixture, with 1 μg each of AP (SEQ ID NO: 21) and CP-10 (which has the sequence 5'-GGGAATTCGATAATCCCATAGACAGC-3')(SEQ ID NO: 9), the nested primer. The nucleotide sequence 5'-GGGAATTC-3' of primer CP-10 (bases 1 through 8 of SEQ ID NO: 9) represent an *Eco* RI restriction site added for cloning purposes. The remaining oligonucleotide sequence encodes amino acids 1-6 of *Cry j* I (AspAsnProIleAspSer) (amino acids 1 through 6 of SEQ ID NO: 1), and is based on the nucleotide sequence determined for the partial *Cry j* I clone JC76.1. The amplified DNA product was purified and precipitated as above, followed by digestion with *Eco* RI and *Xba* I and electrophoresis through a preparative 1% low melt gel. The dominant DNA band was excised and ligated into M13mp19 and pUC19 for sequencing. Again, ligatable

material was recovered from cDNA generated from pollen-derived RNA, and from staminate cone-derived RNA. Two clones, designated pUC19J91A and pUC19J91D, were selected for full-length sequencing. They were subsequently found to have identical sequences.

DNA was sequenced by the dideoxy chain termination method (Sanger et al. *supra*) using a commercially available kit (sequenase kit (U.S. Biochemicals, Cleveland, OH). Both strands were completely sequenced using M13 forward and reverse primers (N.E. Biolabs, Beverly, MA) and internal sequencing primers CP-13 (SEQ ID NO: 10), CP-14 (SEQ ID NO: 11), CP-15 (SEQ ID NO: 12), CP-16 (SEQ ID NO: 13), CP-18 (SEQ ID NO: 15), CP-19 (SEQ ID NO: 16), and CP-20 (SEQ ID NO: 17). CP-13 has the sequence 5'-ATGCCTATGTACATTGC-3' (SEQ ID NO: 10). CP-13 (SEQ ID NO: 10) encodes amino acids 82-87 of *Cry j* I (MetProMetTyrIleAla, amino acids 82 through 87 of SEQ ID NO: 1). CP-14 has the sequence 5'-GCAATGTACATAGGCAT-3' (SEQ ID NO: 11) and corresponds to the non-coding strand sequence of CP-13 SEQ ID NO: 10). CP-15 has the sequence 5'-TCCAATTCTTCTGATGGT-3' ((SEQ ID NO: 12) which encodes amino acids 169-174 of *Cry j* I (SerAsnSerSerAspGly, amino acids 169 through 174 of SEQ ID NO: 1). CP-16 has the sequence 5'-TTTTGTCAATTGAGGAGT-3' (SEQ ID NO: 13) which is the non-coding strand sequence which corresponds to coding strand sequence encoding amino acids 335-340 of *Cry j* I (ThrProGlnLeuThrLys, amino acids 335 through 340 of SEQ ID NO: 1). CP-18 has the sequence 5'-TAGAACTCCAGTCGAAGT-3' (SEQ ID NO: 15) which is the non-coding strand sequence which corresponds to coding strand sequence encoding amino acids 181 through 186 of *Cry j* I (ThrSerThrGlyValThr, amino acids 181 through 186 of SEQ ID NO: 1). CP-19 which has the sequence 5'-TAGCTCTCATTTGGTGC-3' (SEQ ID NO: 16) is the non-coding strand sequence which corresponds to coding strand sequence encoding amino acids 270 through 275 of *Cry j* I (AlaProAsnGluSerTyr, amino acids 270 through 275 of SEQ ID NO: 1). CP-20 has the sequence 5'-TATGCAATTGGTGGGAGT-3' (SEQ ID NO: 17) which is the coding strand sequence for amino acids 251-256 of *Cry j* I

(TyrAlaIleGlyGlySer, amino acids 251 through 256 of SEQ ID NO: 1). The sequenced DNA was found to have the sequence shown in Figures 1a and 1b (SEQ ID NO: 1). This is a composite sequence from the two overlapping clones JC 71.6 and pUC19J91A. The complete cDNA sequence for *Cry j* I is composed of 1312 nucleotides, including 66 nucleotides of 5' untranslated sequence, an open reading frame starting with the codon for an initiating methionine, of 1122 nucleotides, and a 3' untranslated region. There is a consensus polyadenylation signal sequence in the 3' untranslated region 25 nucleotides 5' to the poly A tail. The position of the initiating methionine is confirmed by the presence of in-frame upstream stop codons and by 78% homology with the plant consensus sequence that encompasses the initiating methionine (AAAA<u>AUG</u>GA (bases 62 through 70 of SEQ ID NO: 1) found in *Cry j* I compared with the AACA<u>AUG</u>GC consensus sequence, Lutcke et al. (1987) EMBO J. <u>6</u>: 43-48). The open reading frame encodes a protein of 374 amino acids of which the first 21 amino acids comprise a leader sequence that is cleaved from the mature protein. The amino terminus of the mature protein was identified by comparison with the published NH_2-terminal sequence (Taniai et al. *supra*). The deduced amino acid sequence of the mature protein, comprised of 353 amino acid, has complete sequence identity with the published protein sequence for *Cry j* I (Taniai et al. *supra*, including the first twenty amino acids for the NH_2-terminal and sixteen contiguous internal amino acids. The mature protein also contains five potential N-linked glycosylation sites corresponding to the consensus sequence N-X-S/T.

Example 3

Expression of *Cry j* I

Expression of *Cry j* I will performed as follows. A full-length *Cry j* I cDNA will be generated by polymerase chain reaction (PCR) amplification from first strand cDNA or double-stranded cDNA prepared from mRNA or cloned *Cry j* I with the oligonucleotides CP-10 (SEQ ID NO: 9) and CP-17 (SEQ ID NO: 14). CP-17 has the sequence 5'-

CCTGCAGAAGCTTCATCAACAACGTTTAGA-3'(SEQ ID NO: 14). In CP-17, the sequence 5'-CCTGCAGAAGCTT-3' (bases 1 through 13 of SEQ ID NO: 17) represents *Pst* I and *Hind* III restriction sites added for cloning purposes. The remaining nucleotides of CP-17 (SEQ ID NO: 14) are the non-coding strand sequence corresponding to coding strand sequence that encodes amino acids 350-353 of *Cry j* I (SerLysArgCys, amino acids 350 through 353 of SEQ ID NO: 1) and the translation stop codon. The cDNA sequence will be amplified by 25-30 cycles of denaturation, hybridization and elongation as described herein. The cDNA will be digested with the appropriate restriction enzymes and recovered as described herein and ligated into the expression vector pET11.d (Novagen, Madison, WI; Jameel et al. (1991) J. Virol. 64: 3963-3966.) modified to contain a sequence encoding 6 histidines immediately 3' of the ATG initiation codon and a RNA stabilizing sequence (Skoglund et al. (1990) Gene 88: 1-5) at the 3' end of the cloning site. The histidine (His_6) sequence is added for affinity purification of the recombinant protein *Cry j* I on a Ni^{2+} chelating column according to the method of Hochuli et al (1987) J. Chromatog. 411: 177-184; and Hochuli et al. (1988) BioTech 6: 1321-1325. The recombinant clone will be used to transform *Escherichia coli*, preferably *E. coli* strain BL21-DE3, which harbors a plasmid that has an isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible promoter preceding the gene encoding T7 polymerase according to a method in Sambrook et al., *supra* or other method for transforming *E. coli* cells. Induction with IPTG leads to high levels of T7 polymerase expression, which is necessary for expression of the recombinant protein in pET.11d which has a T7 promoter.

Expression of the recombinant protein will be confirmed in an initial small culture (50 ml). A clone will be grown in media (Brain Heart Infusion Media, Difco, Detroit, MI) containing ampicillin (100 μg/ml), and induced with IPTG (1mM, final concentration) for 2 hrs. The bacteria will be collected by centrifugation, and a crude cell lysate will be prepared by boiling the pellet for 5 minutes in 50 mM tris

HCl, pH 6.8 2 mM EDTA, 1% SDS, 1% β-mercaptoethanol, 10% glycerol, 0.25% bromophenol blue (Studier et al. In: *Methods in Enzymology*, vol. 185, Goeddel, D.V., ed., Academic Press, 1990). Recombinant protein expression will be visualized as a band with the predicted molecular weight on a Coomassie blue-stained SDS-PAGE gel, according to the method in Sambrook et al, *supra*, on which 40 μl of the crude lysate is loaded. Controls will consist of crude lysates from uninduced bacteria containing the plasmid with *Cry j* I, or induced lysates from bacteria carrying plasmids with non-*Cry j* I inserts.

A clone containing the *Cry j* I plasmid will be grown on a large scale for recombinant protein expression and purification. A 2 ml culture of bacteria containing the recombinant plasmid will be grown for 1-2 hr, then streaked onto solid media (e.g. 6 petri plates (100 x 15 mm) with 1.2% agarose in LD medium (Bibco-BRL, Gaithersburg, MD) containing 100 μl/ml ampicillin), grown to confluence, then scraped into 9 L of liquid media (Brain Heart Infusion media, Difco, Detroit, MI) containing ampicillin (100 μg/ml). The culture will be grown until the A_{600} is 1.0, IPTG will be added (1 mM final concentration), and the culture will be grown for an additional 2 hours.

Bacteria will be recovered by centrifugation (7,930 x g, 10 min), and resuspended in 90 ml of 100 mM Na_2HPO_4, 50 mM NaCl containing lysozyme (0.2 mg/ml), protease inhibitors (1mM PMSF [phenylmethyl sulphonyl fluoride], 2 μg/ml Leupeptin, 2 μg/ml Peptstatin A, 1 μg/ml SBTI [soy bean trypsin inhibitor]). After a 30 min incubation on ice, the lysate will be frozen at - 70° C for a minimum of 30 min. Proteins in the inclusion bodies will then be purified by the method of Hecht et al. (1990) Science 249: 884-891. The lysate will be sonicated and the inclusion bodies will be recovered by centrifugation (11,000 x g, 10 min, 4°C). The inclusion bodies will be washed in lysis buffer, without lysozyme, repelleted as above, then solubilized in 6 M Guanidine HCl, 100 mM Na_2HPO_4, pH 8.0, 100 mM β-mercaptoethanol by sonication. Insoluble material will be

removed by centrifugation (11,000 x g, 10 min, 4°C). The pH of the lysate will then be adjusted to pH 8.0, and the lysate applied to an 80 ml Nickel NTA agarose column (Qiagen, Chatsworth, CA) that has been equilibrated with 6 M guanidine HCl, 100 mM Na_2HPO_4, pH 8.0. The column will be sequentially washed with 6 M Guanidine HCl, 100 MM Na_2HPO_4, pH 8.0, then 8 M urea, 100 mM Na_2HPO_4, pH 8.0, and finally 8 M urea, 100 mM sodium acetate, pH 6.3. The column will be washed with each buffer until the flow through has an $A_{280} \leq 0.05$.

The recombinant protein, *Cry j* I, will then be eluted with 8 M urea, 100 mM sodium acetate, pH 4.5, and collected in 10 ml aliquots. The protein concentration of each fraction will be determined by A_{280} and the peak fractions will be pooled. An aliquot of the collected recombinant protein will be analyzed on SDS-PAGE according to the method in Sambrook et al. *supra*. If necessary, the recombinant purified *Cry j* I can be further purified using standard biochemical techniques, such as size exclusion chromatography, reverse phase chromatography or ion exchange chromatography.